softmax求导

softmax:

$$s_i = \frac{e^{a_i}}{\sum_j e^{a_j}}$$

求导: $\frac{\partial s_i}{\partial i} = \frac{\partial\left(\frac{e^{a_i}}{\sum_j e^{a_j}}\right)}{\partial i}$

$$= \frac{\partial\left(1-\frac{\sum_{j\neq i} e^{a_j}}{\sum_j e^{a_j}}\right)}{\partial i}$$

$$= -\sum_{j\neq i} e^{a_j} \cdot \frac{\partial\left(\frac{1}{\sum_j e^{a_j}}\right)}{\partial i}$$

$$= -\cancel{\sum_{j\neq}}(1-e^{a_i}) \frac{-e^{a_i}}{\left(\sum_j e^{a_j}\right)^2}$$

$$= \frac{(1-e^{a_i})\cdot e^{a_i}}{\left(\sum_j e^{a_j}\right)^2}$$

$$= \frac{1-e^{a_i}}{\sum_j e^{a_j}} \cdot \frac{e^{a_i}}{\sum_j e^{a_j}}$$

$$= \frac{\sum_j e^{a_j}-e^{a_i}}{\sum_j e^{a_j}} \cdot s_i$$

$$= (1-s_i)\cdot s_i$$

$\frac{e^i}{(\Sigma e^j)^2}$

cross entropy loss 求导

$$\begin{cases} CE\_loss = -\sum_i t_i \ln y_i \\ y_i = \dfrac{e^i}{\sum_j e^j} = 1 - \dfrac{\sum_{j\neq i} e^j}{\sum_j e^j} \end{cases}$$

其中：t_i 表示真实值（预测第i个，则t_i=1，其余为0）

y_i 表示 softmax 之后的值

~~求导：~~ $\cancel{\dfrac{\partial CE\_loss}{\partial i}} =$

当预测第i个时，$t_i=1$.

$$Loss_i = -\ln y_i$$

对第i个求导：

$$\frac{\partial Loss_i}{\partial i} = \frac{\partial(-\ln y_i)}{\partial i} = \frac{\partial\left(-\ln \frac{e^i}{\sum_j e^j}\right)}{\partial i}$$

$$= -\frac{1}{\frac{e^i}{\sum_j e^j}} \cdot \frac{\partial \frac{e^i}{\sum_j e^j}}{\partial i} = -\frac{\sum_j e^j}{e^i} \frac{\partial\left(1 - \frac{\sum_{j\neq i} e^j}{\sum_j e^j}\right)}{\partial i}$$

$$= -\frac{\sum_j e^j}{e^i}\left(-\sum_{j\neq i} e^j\right) \cdot \frac{\partial\left(\frac{1}{\sum_j e^j}\right)}{\partial i} = \frac{\sum_j e^j \cdot \sum_{j\neq i} e^j}{e^i} \cdot \frac{-e^i}{(\sum_j e^j)^2}$$

$$= -\frac{\sum_{j\neq i} e^j}{\sum_j e^j} = -\left(1 - \frac{e^i}{\sum_j e^j}\right) = -(1-y_i) = y_i - 1$$

参考链接

1. https://blog.csdn.net/bitcarmanlee/article/details/105619286
2. https://blog.csdn.net/grllery/article/details/97788745
3. https://blog.csdn.net/bitcarmanlee/article/details/82320853